Speech Processing 2012/13

2nd Test

May 2nd 2013

Please identify this form with your name and student number in the reserved space. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. Wrong answers to these questions will be penalized. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

1. What is the name of the PCM variant which has been standardized most efficiently at several bit rates?
2. Some narrowband coders that take advantage of the fact that in formant regions noise can be partially masked by the signal.
   (a) Name three of these coders.
   (b) Do they achieve this by (A) increasing or (B) reducing the bandwidth of the roots of the short-term predictor used in the stochastic codebook search?
   (c) Write the transfer function of a short-term predictor with modified formant bandwidths.
3. A given subband speech coder splits the bandwith (100-3200 Hz) into 5 bands using QMF filters. Justify how this coder may achieve a bit rate close to 16kbps, by indicating the 5 frequency bands and corresponding sampling rates, and assigning a different number of bits to each band. 1) 0-1600, 1600-3200 2) 0-800, 800-1600 3) 0-400, 400-800 4) 0-200, 200-400 5) 0-200, 200-400
4. Classify as True (T) or False (F)
   (a) In CELP coders, the highest contribution to the total bit rate is typically related to the short-term predictor.
   (b) Phonetic vocoders may use pitch-dependent codebooks of LPC coefficients in their synthesis filter.
   (c) A double prediction residual exhibits harmonics at least in the lower frequency bands.
   (d) Vector quantization achieves a better approximation of the lower distortion bound than instantaneous quantization.
   (e) In RPE coders, the pulse grid index is typically quantized with 400 bps.
5. Classify as True (T) or False (F)
   (a) In Portuguese, the general stress rule applies to words ending in vowels and words ending in consonants.
   (b) Prosodic phrasing modules based on CARTs take into account the number of words that can be pronounced without pausing for breath, by asking questions about the PoS of each word and the next one.
   (c) In formant synthesizers one needs to fine tune the parameters that control the transitions between phones.
   (d) In the PSOLA method, pitch and duration modifications can be done independently of each other.
   (e) Statistical parametric synthesis models can be characterized by over smoothing.
   (f) The performance of a neural network such as a multilayer perceptron applied to grapheme-to-phone conversion depends on the number of graphemes considered as left and right context.

6. Pretend you want to design tests to assess synthesizers in terms of intelligibility.

   (a) Name two tests that are commonly used for that purpose.

   (b) Give examples of the text materials that could be read for each of these tests (if you are a non-native, the examples may be in English).

7. Write two pronunciations for the following three foreign company names. The first pronunciation should be the one typically heard in Portuguese media. The second pronunciation is the one obtained using the grapheme-to-phone rules for the common lexicon in European Portuguese. Write also the most used pronunciation of the acronym in (d). All transcriptions should use only the phonetic SAMPA symbols of this language.

   (a) Apple

   (b) Verizon Communications

   (c) China Insurance

   (d) TMN

8. Pretend you have developed a limited domain synthesizer, for a talking calendar, which can only synthesize the names of the days in the week in European Portuguese, plus some welcoming word sequences: *calendário falado, hoje é, domingo, segunda, terça, quarta, quinta, sexta, sábado, feira, obrigado pela sua preferência*.

   (a) Indicate which phone sequences you would select from which words to synthesize by concatenation a new word - *feriado* - using the fewest sequences. Joint points do not occur in phone transitions. If you miss any phone sequence indicate which one. Use # as a symbol for silence.

   (b) The prosody generation module computed a target duration of 80ms for the fricative sound of the new word. The left sub-unit has duration values of 44 and 50ms for the two parts. The right sub-unit has duration values of 66 and 22ms for the two parts. Indicate the target durations for the two sub-units of the fricative.

9. Using the following syntax
   a $\rightarrow$ b / c .. d
   describe the simplified rules for grapheme *n*, where *c* and *d* may be graphemes (e.g.: a, b, etc.), classes of graphemes (vowels, consonants, etc.), the word boundary (#), or any grapheme (*). You may use the symbol *0* to mark phonemic nulls, and the symbols | and ( ) to mark disjunction between several graphemes (e.g.: a | b | c). The rules are applied in order, until one matches the context and, in this case, the following rules are not applied. The "simpliflied" rules do not need to contemplate compound words, or words of foreign origin. Here are some examples of words illustrating the different pronunciations of this grapheme (in alphabetical order):
   ano, canto, conferência, forno, íman, não, nunca, unha

Test 2 - Answers

Name:	
Number:	
Group Number:	

1 (1.0 val.)

2 (3.2 val.)

a	
b	
c	

3 (2.0 val.)

4. (1.5 val.) Indicate T or F:

a	b	c	d	e

5. (1.8 val.) Indicate T or F:

a	b	c	d	e	f

6 (2.0 val.)

a	
b	

7. (2.8 val.)

Apple		
Verizon Communications		
China Insurance		
TMN		-

8. (3.7 val.)

a	
b	

9 (3.0 val.)